

AN EMPIRICALLY DERIVED FIGURE OF MERIT FOR THE QUALITY OF
OVERALL TASK PERFORMANCE

Moira LeMay
Associate Professor
Department of Psychology
Montclair State College
Upper Montclair, NJ 07043

The need to develop an operationally relevant figure of merit for the quality of performance of a complex system such as an airplane cockpit stems from a hypothesized dissociation between measures of performance and those of workload (ref. 1). At moderate workload levels, increasing task demands generally leave performance unaffected if operators have sufficient spare capacity. Overload or underload conditions both lead to task performance problems. Reasons for concern with workload measurement include prediction and prevention of system failure and definition of "optimal" workload level which produces smooth system performance. These require a figure of merit to measure overall quality of performance in order to gauge the effect of workload and to signal impending system failure.

Performance can be measured in terms of time, errors, or a combination of these. In most tasks performed by expert operators, errors are relatively rare and often corrected in time to avoid consequences. Moreover, perfect performance is seldom necessary to accomplish a particular task, e.g., a pilot does not need to follow an assigned flight path with zero error and, indeed, does not (ref. 2), but only keeps the craft within certain bounds. Moreover, how well an expert performs a complex task consisting of a series of discrete cognitive tasks superimposed on a continuous task, such as flying an airplane, does not depend on how "well" each discrete task is performed, but on their smooth sequencing. This makes amount of time spent on each subtask of paramount importance in measuring overall performance, since smooth sequencing requires a minimum amount of time spent on each task. Quality consists in getting tasks done within a critical time interval while maintaining acceptable continuous task performance. Thus, a figure of merit for overall quality of performance should be primarily a measure of time to perform discrete subtasks combined with a measure of basic vehicle control.

Acceptable level of performance produces a safe landing. Since most landings are safe, average pilot performance is acceptable. Therefore, it is proposed that deviation from average pilot performance be used as a standard for both continuous and discrete pilot tasks. This allows for a reasonable, rather than absolute, standard and makes it possible to express performance in terms of standard units.

Thus, the proposed figure of merit requires doing a task analysis on a series of performances, or runs, of a particular task, listing each discrete task and its associated time, and calculating the mean and standard deviation of these times, along with the mean and standard deviation of tracking error for the whole task. Since most discrete tasks are cognitive, they are best timed if the pilot indicates the start and end of each task with a keystroke, although other means may be used. Time for each task receives a single standard unit (one unit is one standard deviation above or below the mean). Such a unit can then be combined additively with other standard units for

very different tasks. Since a simple addition of equally weighted standard units will not reflect differences in relative contribution of various tasks, the standard unit for each task must be multiplied by the mean time for that task and the tasks added together to obtain a figure of merit. This can be calculated for each new run, which then receives its individual score. Such scores are automatically given in terms of the norm for the whole task.

A set of simulator data on 30 runs of a landing task (ref. 3) has been obtained and a figure of merit will be calculated for each run. On half of the runs, communications with ATC were done with conventional voice radio and on the other half, they were done with a data link ATC message exchange unit. An example of a time line showing timing of discrete and continuous tasks for two runs is illustrated in Figure 1. The figure of merit will be compared for voice and data link, so that the impact of this technology on total crew performance (not just communication performance) can be assessed. The effect of data link communication on other cockpit tasks will also be considered. In this way, it is hoped that the usefulness of the proposed figure of merit as a measure of the impact of technology introduction can be demonstrated.

In addition, a complex task consisting of a number of subtasks is currently being developed in order to study the impact of very low task demands. It will be possible to calculate a more accurate figure of merit for each run of this task, since data collection can be planned in advance. It is hoped that such data will also show appropriate variations in the figure of merit, and the conditions of the study will be reflected in it.

The figure of merit thus developed should then serve as a measure of overall task performance against which variations in workload can be assessed. It should be used to explore the ways in which such task factors as boredom, overload, and their alternation, as measured by traditional workload measures such as physiological variations and subjective impressions, may influence task performance.

References

1. Lysaght, R.J., Hill, S.G., Dick, A.O., Plamondon, B.D., Wherry, R.J., Zaklad, A.L., & Bittner, A.C. Operator workload: comprehensive review and evaluation of operator workload methodologies. Analytics Tech Rept 2075-3. ARI contract No. MDA 903-86-C-0384 Item no. 002AG (revised) 1988.
2. Brietson, C.A., Ciavarelli, A.P., and Wulfeck, J.W. Operational measures of aircraft carrier landing system performance. Human Factors, 11, 281-290, 1969.
3. Waller, M.C. and Lohr, G.W. A Piloted Simulation Study of Data Link ATC Message Exchange. NASA TP-2859, 1989.

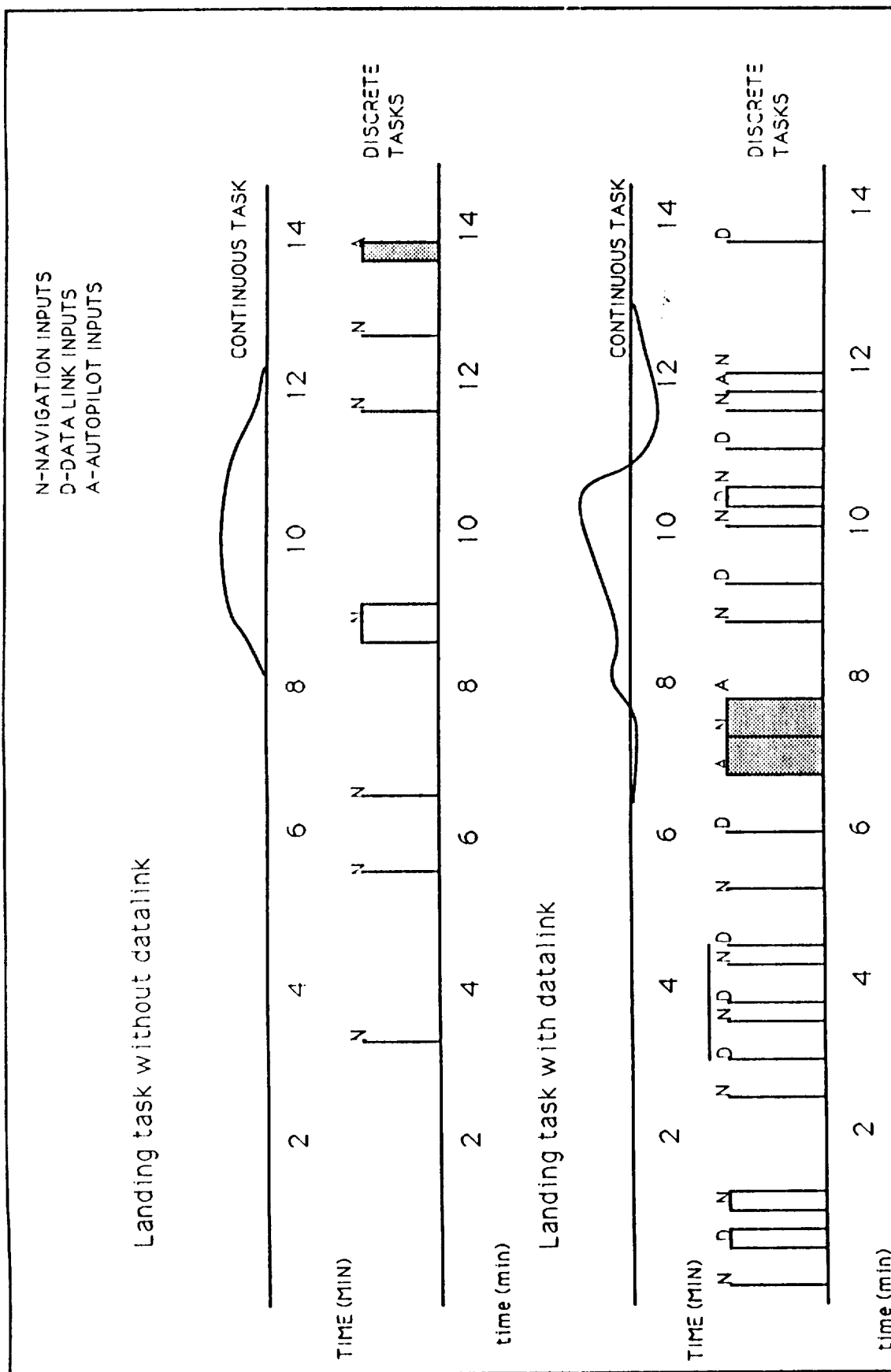


FIGURE 1. Distance from flight path (continuous task) and occurrence of discrete tasks in a simulated landing task.